



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/823,617	04/14/2004	Randall J. Calistri-Yeh	55653-016	5090
7590 McDermott, Will & Emery 600 13th Street, N.W. Washington, DC 20005-3096		09/18/2007	EXAMINER PANNALA, SATHYANARAYAN R	
			ART UNIT 2164	PAPER NUMBER
			MAIL DATE 09/18/2007	DELIVERY MODE PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

80

Interview Summary	Application No.	Applicant(s)	
	10/823,617	CALISTRI-YEH ET AL.	
	Examiner	Art Unit	
	Sathyanarayan Pannala	2164	

All participants (applicant, applicant's representative, PTO personnel):

- (1) Sathyanarayan Pannala. (3) _____.
- (2) Nichlos Chen. (4) _____.

Date of Interview: 12 September 2007.

Type: a) ☒ Telephonic b) ☐ Video Conference
 c) ☐ Personal [copy given to: 1) ☐ applicant 2) ☐ applicant's representative]

Exhibit shown or demonstration conducted: d) ☐ Yes e) ☐ No.
 If Yes, brief description: _____.

Claim(s) discussed: 41,46,50,63,64,69 and 70.

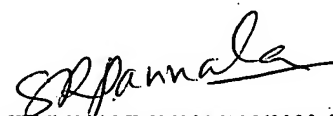
Identification of prior art discussed: N/A.

Agreement with respect to the claims f) ☐ was reached. g) ☒ was not reached. h) ☐ N/A.

Substance of Interview including description of the general nature of what was agreed to if an agreement was reached, or any other comments: Applicant discussed proposed amendment to overcome rejection under 35 U.S.C. 101. Examiner explained how to overcome the rejection. No agreement was reached.

(A fuller description, if necessary, and a copy of the amendments which the examiner agreed would render the claims allowable, if available, must be attached. Also, where no copy of the amendments that would render the claims allowable is available, a summary thereof must be attached.)

THE FORMAL WRITTEN REPLY TO THE LAST OFFICE ACTION MUST INCLUDE THE SUBSTANCE OF THE INTERVIEW. (See MPEP Section 713.04). If a reply to the last Office action has already been filed, APPLICANT IS GIVEN A NON-EXTENDABLE PERIOD OF THE LONGER OF ONE MONTH OR THIRTY DAYS FROM THIS INTERVIEW DATE, OR THE MAILING DATE OF THIS INTERVIEW SUMMARY FORM, WHICHEVER IS LATER, TO FILE A STATEMENT OF THE SUBSTANCE OF THE INTERVIEW. See Summary of Record of Interview requirements on reverse side or on attached sheet.


SATHYANARAYAN PANNALA
PRIMARY EXAMINER

Examiner Note: You must sign this form unless it is an Attachment to a signed Office action.

Examiner's signature, if required

Summary of Record of Interview Requirements

Manual of Patent Examining Procedure (MPEP), Section 713.04, Substance of Interview Must be Made of Record

A complete written statement as to the substance of any face-to-face, video conference, or telephone interview with regard to an application must be made of record in the application whether or not an agreement with the examiner was reached at the interview.

Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews Paragraph (b)

In every instance where reconsideration is requested in view of an interview with an examiner, a complete written statement of the reasons presented at the interview as warranting favorable action must be filed by the applicant. An interview does not remove the necessity for reply to Office action as specified in §§ 1.111, 1.135. (35 U.S.C. 132)

37 CFR §1.2 Business to be transacted in writing.

All business with the Patent or Trademark Office should be transacted in writing. The personal attendance of applicants or their attorneys or agents at the Patent and Trademark Office is unnecessary. The action of the Patent and Trademark Office will be based exclusively on the written record in the Office. No attention will be paid to any alleged oral promise, stipulation, or understanding in relation to which there is disagreement or doubt.

The action of the Patent and Trademark Office cannot be based exclusively on the written record in the Office if that record is itself incomplete through the failure to record the substance of interviews.

It is the responsibility of the applicant or the attorney or agent to make the substance of an interview of record in the application file, unless the examiner indicates he or she will do so. It is the examiner's responsibility to see that such a record is made and to correct material inaccuracies which bear directly on the question of patentability.

Examiners must complete an Interview Summary Form for each interview held where a matter of substance has been discussed during the interview by checking the appropriate boxes and filling in the blanks. Discussions regarding only procedural matters, directed solely to restriction requirements for which interview recordation is otherwise provided for in Section 812.01 of the Manual of Patent Examining Procedure, or pointing out typographical errors or unreadable script in Office actions or the like, are excluded from the interview recordation procedures below. Where the substance of an interview is completely recorded in an Examiners Amendment, no separate Interview Summary Record is required.

The Interview Summary Form shall be given an appropriate Paper No., placed in the right hand portion of the file, and listed on the "Contents" section of the file wrapper. In a personal interview, a duplicate of the Form is given to the applicant (or attorney or agent) at the conclusion of the interview. In the case of a telephone or video-conference interview, the copy is mailed to the applicant's correspondence address either with or prior to the next official communication. If additional correspondence from the examiner is not likely before an allowance or if other circumstances dictate, the Form should be mailed promptly after the interview rather than with the next official communication.

The Form provides for recordation of the following information:

- Application Number (Series Code and Serial Number)
- Name of applicant
- Name of examiner
- Date of interview
- Type of interview (telephonic, video-conference, or personal)
- Name of participant(s) (applicant, attorney or agent, examiner, other PTO personnel, etc.)
- An indication whether or not an exhibit was shown or a demonstration conducted
- An identification of the specific prior art discussed
- An indication whether an agreement was reached and if so, a description of the general nature of the agreement (may be by attachment of a copy of amendments or claims agreed as being allowable). Note: Agreement as to allowability is tentative and does not restrict further action by the examiner to the contrary.
- The signature of the examiner who conducted the interview (if Form is not an attachment to a signed Office action)

It is desirable that the examiner orally remind the applicant of his or her obligation to record the substance of the interview of each case. It should be noted, however, that the Interview Summary Form will not normally be considered a complete and proper recordation of the interview unless it includes, or is supplemented by the applicant or the examiner to include, all of the applicable items required below concerning the substance of the interview.

A complete and proper recordation of the substance of any interview should include at least the following applicable items:

- 1) A brief description of the nature of any exhibit shown or any demonstration conducted,
- 2) an identification of the claims discussed,
- 3) an identification of the specific prior art discussed,
- 4) an identification of the principal proposed amendments of a substantive nature discussed, unless these are already described on the Interview Summary Form completed by the Examiner,
- 5) a brief identification of the general thrust of the principal arguments presented to the examiner,
(The identification of arguments need not be lengthy or elaborate. A verbatim or highly detailed description of the arguments is not required. The identification of the arguments is sufficient if the general nature or thrust of the principal arguments made to the examiner can be understood in the context of the application file. Of course, the applicant may desire to emphasize and fully describe those arguments which he or she feels were or might be persuasive to the examiner.)
- 6) a general indication of any other pertinent matters discussed, and
- 7) if appropriate, the general results or outcome of the interview unless already described in the Interview Summary Form completed by the examiner.

Examiners are expected to carefully review the applicant's record of the substance of an interview. If the record is not complete and accurate, the examiner will give the applicant an extendable one month time period to correct the record.

Examiner to Check for Accuracy

If the claims are allowable for other reasons of record, the examiner should send a letter setting forth the examiner's version of the statement attributed to him or her. If the record is complete and accurate, the examiner should place the indication, "Interview Record OK" on the paper recording the substance of the interview along with the date and the examiner's initials.

McDermott Will & Emery

Boston Brussels Chicago Düsseldorf London Los Angeles Miami Munich
New York Orange County Rome San Diego Silicon Valley Washington, D.C.

Strategic alliance with MWE China Law Offices (Shanghai)

FACSIMILE

Date: September 10, 2007

Time Sent:

To:	Company:	Facsimile No:	Telephone No:
Examiner Sathyanarayan R. Pannala	USPTO	1.571.273.4115	
From:	W. Nicholas Chen	Direct Phone:	650.813.5092
E-Mail:	nchen@mwe.com	Direct Fax:	650.813.5100
Sent By:	Patricia A. Balero	Direct Phone:	650.813.4608
Client/Matter/Tkpr:	055653-0016/09341	Original to Follow by Mail:	No
		Number of Pages, Including Cover:	16
Re:	Interview Agenda for Tele-Conference		

Message:

The information contained in this facsimile message is legally privileged and confidential information intended only for the use of the individual or entity named above. If the reader of this message is not the intended recipient, you are hereby notified that any dissemination, distribution, or copy of this facsimile is strictly prohibited. If you have received this facsimile in error, please notify us immediately by telephone and return the original message to us at the below address by mail. Thank you.

IF YOU DO NOT RECEIVE ALL OF THE PAGES, PLEASE CALL AS SOON AS POSSIBLE.

Main Facsimile: 202.756.8087 Facsimile Operator: 202.756.8090

U.S. practice conducted through McDermott Will & Emery LLP.
600 Thirteenth Street, N.W. Washington, D.C. 20005-3096

Telephone: 202.756.8000

WDC99 1459510-1.055653.0016

PAGE 1/16 * RCVD AT 9/10/2007 7:57:46 PM [Eastern Daylight Time] * SVR:USPTO-EFAX-6/1 * DNIS:2734115 * CSID: * DURATION (mm-ss):02-28

PTOL-413A (07-07)
Approved for use through 09/30/2007. OMB 0851-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Applicant Initiated Interview Request Form

Application No.: 10/823,617 First Named Applicant: Randall J. CALISTRI-YEH, et al.
Examiner: S. R. Pannala Art Unit: 2164 Status of Application: Pending

Tentative Participants:

(1) Wei-Chen Nicholas Chen (2) S.R. Pannala
(3) _____ (4) _____

Proposed Date of Interview: 09/12/2007 Proposed Time: 4:00 EST. (AM/PM) (PM)

Type of Interview Requested:

(1) ☒ Telephonic (2) ☐ Personal (3) ☐ Video Conference

Exhibit To Be Shown or Demonstrated: ☒ YES ☐ NO

If yes, provide brief description: Proposed claim amendment

Issues To Be Discussed

Issues (Rej., Obj., etc)	Claims/ Fig. #s	Prior Art	Discussed	Agreed	Not Agreed
(1) <u>\$101</u>	<u>41-52, 63</u> <u>64, 69 &</u> <u>70</u>	_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(2) _____	_____	_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(3) _____	_____	_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(4) _____	_____	_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Continuation Sheet Attached			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Brief Description of Arguments to be Presented:

See the attached agenda

An interview was conducted on the above-identified application on _____

NOTE: This form should be completed by applicant and submitted to the examiner in advance of the interview (see MPEP § 713.01).

This application will not be delayed from issue because of applicant's failure to submit a written record of this interview. Therefore, applicant is advised to file a statement of the substance of this interview (37 CFR 1.133(b)) as soon as possible.

Wei-Chen Nicholas Chen
Applicant/Applicant's Representative Signature

Examiner/SPE Signature

Wei-Chen Nicholas Chen
Typed/Printed Name of Applicant or Representative

56,665
Registration Number, if applicable

This collection of information is required by 37 CFR 1.133. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 123 and 37 CFR 1.11 and 1.14. This collection is estimated to take 21 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Docket No.: 055653-0016

PATENT**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of	:	Customer Number: 20277
Randall J. CALISTRI-YEH, et al.	:	Confirmation Number: 5090
Serial No.: 10/823,617	:	Group Art Unit: 2164
Filed: April 14, 2004	:	Examiner: Sathyanaraya R. Pannala
For: CONSTRUCTION OF TRAINABLE SEMANTIC VECTORS AND CLUSTERING CLASSIFICATION, AND SEARCHING USING TRAINABLE SEMANTIC VECTORS		

INTERVIEW AGENDA

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Per the Examiner's request, this Agenda is submitted for the upcoming telephone interview to be held on September 12, 2007 at 4:00 pm EST.

The Examiner is thanked for the favorable indication that all pending claims are in condition for allowance if outstanding rejections under 35 U.S.C. §101 are overcome and claim objections are addressed.

It is respectfully submitted that the rejections under 35 U.S.C. §101 are overcome and the claim objections are addressed in view of the remarks and proposed amendment submitted herewith. Alternatively, the Examiner's guidance and thoughts on permissible amendments are respectfully solicited.

It is submitted that the claims are not directed solely to mere ideas, laws of nature, or natural phenomena. Each of the claims falls squarely into one of the classes of subject matter permitted by

Serial No.: 10/823,617

35 U.S.C. § 101, that is to say process or machine, respectively. Independent claims 63 and 64, for example, are tied to a data processing system (machine), and independent claims 41, 46 and 50 are directed to a series of computer-executed steps (process). Independent claims 69 and 70 recite a tangible machine-readable medium, in conformity with *In re Beauregard*, 53 F.3d 1583, 35 USPQ2d 1383 (Fed. Cir. 1995). According to the Beauregard decision, computer programs embodied in a tangible medium, such as floppy diskettes, are statutory subject matter under 35 U.S.C. §101.

In addition, these claims are more than “a computer” that “solely calculates a mathematical formula” or “a computer disk that solely stores a mathematical formula. Rather, the claimed process and system are not mere abstract concept or mathematical formula. Rather, the process and system provide useful, concrete and tangible results, in classifying a file or document relative to a predetermined number of categories. For instance, claims 41, 46 and 50 are directed to a method of classifying datasets relative to a predetermined categories, which is a key element for technologies like search engines or databases in identifying and/or retrieving requested files or documents. Techniques described in claim 41 uniquely describes determining trainable semantic vectors for each category using sample datasets, and advantageously classifies datasets according to a specific attribute (the trainable semantic vector) of the datasets and the categories, which improves a machine’s efficiency in identifying or retrieving data, and differentiating between unrelated data. Documents or files classified in the same category are likely to be related to one another and may be retrieved together, while documents or datasets in different categories are unlikely to be related. Accordingly, the claims describe concepts that create a “useful, non-abstract result” analogous to the method of adding a data field with information on long distance providers, which the Federal Circuit found to be a “useful, non-abstract result that facilitates differential billing of long-distance

Serial No.: 10/823,617

calls," which "fall[s] comfortably within the broad scope of patentable subject matter under §101." Emphasis added. *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352, 50 USPQ2d 1447 (Fed. Cir. 1999).

For reasons outlined above, it is submitted that all of the rejections under 35 U.S.C. § 101 should be withdrawn.

Incidentally, Applicants note that paragraph 5 of the office action asserted that claims 63-64 and 69-70 "are, at best, functional descriptive material per se." However, according to MPEP 2173.05(g), "[t]here is nothing inherently wrong with defining some part of an invention in functional terms. Functional language does not, in and of itself, render a claim inappropriate. A functional limitation must be evaluated and considered, just like any other limitation of the claim." It is respectfully requested that the rejection of claims 63-64 and 69-70 is overcome.

Furthermore, the Office Action asserted that the term "machine-executed" in the claims does not have a corresponding definition in the specification. By this proposed amendment, the term "machine-executed" is amended to "computer-executed." The specification clearly describes "a computer system upon which an embodiment of the invention may be implemented." See page 10, lines 9-10 and Fig. 1.

Serial No.: 10/823,617

PROPOSED AMENDMENTS TO THE CLAIMS:

Claims 1-40 (Cancelled)

(41) (Currently Amended) A method of classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, the method comprising the ~~machine~~computer-executed steps:

constructing a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

constructing a trainable semantic vector for each category based on the trainable semantic vectors for the sample datasets;

receiving a new dataset;

constructing a trainable semantic vector for the new dataset;

determining a distance between the trainable semantic vector for the new dataset and the trainable semantic vector of each category; and

classifying the new dataset within the category whose trainable semantic vector has the shortest distance to the trainable semantic vector of the new dataset;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

Serial No.: 10/823,617

determining the significance of each data point with respect to the predetermined categories;
constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and
combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

42. (Currently Amended) The method of Claim ~~[[41]]41~~, wherein the datasets correspond to documents.

43. (Currently Amended) The method of Claim ~~[[41]]41~~, wherein the datasets correspond to email messages and the categories correspond to frequently asked questions with substantially static responses.

44. (Original) The method of Claim 41, further comprising the steps:
detecting when a prescribed number of new datasets has been classified; and
updating the trainable semantic vectors for each of the categories.

45. (Original) The method of Claim 44, wherein the step of updating comprises the step of re-constructing trainable semantic vectors for each category based on the trainable semantic vectors for the sample datasets and the trainable semantic vectors for the new datasets added to each category.

46. (Currently Amended) A method of classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, the method comprising the ~~machine~~computer-executed steps:

Serial No.: 10/823,617

constructing a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

receiving a new dataset;

constructing a trainable semantic vector for the new dataset;

identifying a select number of sample datasets whose trainable semantic vectors are closest in distance to the trainable semantic vector for the new dataset; and

classifying the new dataset in the category containing the greatest number of the select sample datasets;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

Serial No.: 10/823,617

47. (Currently Amended) The method of Claim ~~[[46]]46~~, wherein the datasets correspond to documents.

48. (Currently Amended) The method of Claim ~~[[46]]46~~, wherein the datasets correspond to email messages and the categories correspond to frequently asked questions with substantially static responses. ~~[[49.]]~~

49. (Original) The method of Claim 46, further comprising the steps:
detecting when a prescribed number of new datasets has been classified; and
adding the new datasets to the set of sample datasets.

50. (Currently Amended) A method of classifying new datasets within a predetermined number of categories, the method comprising the ~~machine~~computer-executed steps:

receiving a new dataset;

constructing a trainable semantic vector for the new dataset, where the dimensions of the trainable semantic vector correspond to the predetermined number of categories;

classifying the dataset in the category whose corresponding dimension in the trainable semantic vector has the largest value;

wherein:

the new data set includes one or more data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for the new dataset is constructed by performing the steps of:

for each data point within the new dataset, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

Serial No.: 10/823,617

determining the significance of each data point with respect to the predetermined categories;
constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and
combining the trainable semantic vector for each data point to form the semantic vector of the new dataset.

51. (Currently Amended) The method of Claim ~~[[50]]50~~, wherein the datasets correspond to documents.

52. (Currently Amended) The method of Claim ~~[[50]]50~~, wherein the datasets correspond to email messages and the categories correspond to frequently asked questions with substantially static responses.

Claims 53-62 (Cancelled)

63. (Previously Amended) A system for classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, the system comprising:

a computer configured to:

construct a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

construct a trainable semantic vector for each category based on the trainable semantic vectors for the sample datasets;

receive a new dataset;

construct a trainable semantic vector for the new dataset;

Serial No.: 10/823,617

determine a distance between the trainable semantic vector for the new dataset and the trainable semantic vector of each category; and

classify the new dataset within the category whose trainable semantic vector has the shortest distance to the trainable semantic vector of the new dataset;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

64. (Previously Amended) A system for classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, the system comprising:

a computer configured to:

Serial No.: 10/823,617

construct a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

receive a new dataset;

construct a trainable semantic vector for the new dataset;

identify a select number of sample datasets whose trainable semantic vectors are closest in distance to the trainable semantic vector for the new dataset; and

classify the new dataset in the category containing the greatest number of the select sample datasets;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories; and

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

Serial No.: 10/823,617

Claims 65-68 (Cancelled)

69. (Currently amended) A computer-readable medium carrying one or more sequences of instructions for classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the machinecomputer-executed steps of:

constructing a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

constructing a trainable semantic vector for each category based on the trainable semantic vectors for the sample datasets;

receiving a new dataset;

constructing a trainable semantic vector for the new dataset;

determining a distance between the trainable semantic vector for the new dataset and the trainable semantic vector of each category; and

classifying the new dataset within the category whose trainable semantic vector has the shortest distance to the trainable semantic vector of the new dataset;

wherein:

the new data set or each of the sample data sets includes at least one data point;

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

Serial No.: 10/823,617

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

70. (Previously Amended) A computer-readable medium carrying one or more sequences of instructions for classifying new datasets within a predetermined number of categories based on assignment of a plurality of sample datasets to each category, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the steps of:

constructing a trainable semantic vector for each sample dataset relative to the predetermined categories in a multi-dimensional semantic space;

receiving a new dataset;

constructing a trainable semantic vector for the new dataset;

identifying a select number of select datasets whose trainable semantic vectors are closest in distance to the trainable semantic vector for the new dataset; and

classifying the new dataset in the category containing the greatest number of the select dataset;

wherein:

the new data set or each of the sample data sets includes at least one data point;

Serial No.: 10/823,617

each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string; and

the trainable semantic vector for each sample data set or the new dataset is constructed by performing the steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories;

constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and

combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the sample dataset or the new dataset.

Claims 71-78 (Cancelled)

Serial No.: 10/823,617

Respectfully submitted,

McDERMOTT WILL & EMERY LLP

A handwritten signature in black ink, appearing to read "Wei-Chen Chen". The signature is fluid and cursive, with the first name "Wei-Chen" and the last name "Chen" clearly distinguishable.

Wei-Chen Nicholas Chen

Registration No. 56,665

650-813-5092